Comment on Wibral et al. (2013): Measuring Information-Transfer Delays

Jakob Runge

Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany

Department of Physics, Humboldt University, 12489 Berlin, Germany

May 1, 2013

In their article Wibral et al. (2013), the authors propose a measure of interaction delays rooted in an information-theoretic framework. Their measure, named TE_{SPO} where SPO stands for *self-prediction optimality*, is a time-delayed extension to transfer entropy that becomes maximal only at the actual interaction delay as proven in their paper.

Wibral et al. only considered the bivariate coupling case. Prior to their publication, in Runge et al. (2012b,a) the theory of detecting and quantifying causal interactions and their strength for the general multivariate case is discussed. However, Wibral et al. present an interesting complementary approach in that they propose to determine the interaction delay based on the reconstructed (vector-valued) states rather than the (scalar) observations of the complex systems under study.

Wibral et al. contrast their measure with a similar information-theoretic measure, the momentary information transfer (MIT), that was introduced in Pompe and Runge (2011). Wibral et al. write that "*A major conceptual difference between the Pompe and Runge study and ours is that no formal proof of the maximality of their functional MIT at the correct interaction delay is given, and as we argue below cannot be given.*" Rather than providing a proof that the maximality of MIT cannot be given, they construct a model example for which they find the MIT unable to serve the purpose of inferring the interaction delay. But, as shown below, the reasoning in their "maximality"-proof equally applies to the MIT and can, thus, not be used to disregard MIT. Rather, their example model seems to not fulfill the assumptions implicitly used in their own proof.

The "formal proof of the maximality" of MIT was actually given in subsequent works in Runge et al. (2012b,a). The latter works further developed the original idea of Pompe and Runge (2011) in that a two-step approach is proposed. In the first step the causal graph (conditional independence graph) is reconstructed (Runge et al., 2012b) using the well-established framework of graphical models, where the property to capture the correct causal interaction delays is a trivial consequence of separation properties in the graph. In Runge et al. (2012b) also the underlying assumptions for such an inference are

given. In the second step the MIT is used as a measure of the coupling strength solely of the causal links, i.e., the inferred interaction delays, in this graph (Runge et al., 2012a).

Wibral at al. also have overseen that their estimator of conditional mutual information given by their Eq. (19) has already been developed in Frenzel and Pompe (2007). The latter work also discussed the inference of interaction delays.

Proof that MIT is able to infer the correct interaction delay

Wibral et al. contrast their measure with the above-mentioned information-theoretic quantity MIT, that has been thoroughly studied in Runge et al. (2012a) and was introduced in the older publication in Pompe and Runge (2011). It is defined as

$$I^{\mathrm{MIT}}_{X\to Y}(\tau) \equiv I(X_{t-\tau}; Y_t | \mathcal{P}_{Y_t} \setminus \{X_{t-\tau}\}, \mathcal{P}_{X_{t-\tau}}), \tag{1}$$

i.e., additionally to ITY, the parents of X are included in the condition. Why this is reasonable is extensively discussed in Runge et al. (2012a). The discussion in Runge et al. (2012a) focused on the point of what a reasonable measure of the interaction strength is, given that *the interaction delays are already known since they have been inferred using the algorithm introduced in Runge et al. (2012b).*

Wibral et al. prove that their TE_{SPO} is maximal only for the interaction delay. To this end they use the Markov properties of the process and information theoretic properties of the conditional mutual information like the data processing inequality and the chain rule (Cover and Thomas, 2006).

To show that this reasoning equally applies to MIT, we follow their proof and add the modifications for MIT in red, i.e., the additional condition on the parents of the lagged X, written as $\mathcal{P}_{X_{t-\delta-\xi}}$. Thus, their proof is obtained by leaving out the red variables. For a multivariate process (X, Y) with conditional independencies given by the graph shown in their Figure 2, we apply the chain rule for conditional mutual information twice ($\xi \neq 0$):

$$\begin{aligned} & I(Y_t; (X_{t-\delta-\xi}, X_{t-\delta}) | Y_{t-1}, \mathcal{P}_{X_{t-\delta-\xi}}) \\ &= I(Y_t; X_{t-\delta-\xi} | Y_{t-1}, \mathcal{P}_{X_{t-\delta-\xi}}) + \underbrace{I(Y_t; X_{t-\delta} | Y_{t-1}, X_{t-\delta-\xi}, \mathcal{P}_{X_{t-\delta-\xi}})}_{\geq 0 \quad \text{due to non-negativity of CMI}} \end{aligned} \tag{2}$$

$$= I(Y_t; X_{t-\delta} | Y_{t-1}, \mathcal{P}_{X_{t-\delta-\xi}}) + \underbrace{I(Y_t; X_{t-\delta-\xi} | Y_{t-1}, X_{t-\delta}, \mathcal{P}_{X_{t-\delta-\xi}})}_{=0 \text{ due to } d\text{-separation implying conditional independence}} \tag{3}$$

Thus,

$$I(Y_t; X_{t-\delta-\xi} | Y_{t-1}, \mathcal{P}_{X_{t-\delta-\xi}}) \leq I(Y_t; X_{t-\delta} | Y_{t-1}, \mathcal{P}_{X_{t-\delta-\xi}}), \tag{4}$$

showing that Wibral's theorem applies to both their TE_{SPO} and MIT. If their example model contradicts this property, it simply means that their example model does not fulfill the very basic assumptions used in this proof. Possibly the notion of *d*-separation. Unfortunately, Wibral et al. do not give a derivation of their analytical computation of MIT for their model. In Runge et al. (2012b) and Runge et al. (2012a) we refer to the condition (S) in Eichler (2012) that a process must fulfill to guarantee that *d*-separation in the causal graph implies conditional independence in the process. The model actually features a *deterministic* coupling from X to Y and randomness only in X. If even this remaining randomness is set to very small values, only then does the "inability" of MIT appear. Apart from the question how realistic a purely deterministic dependence is in real applications, generally, information theory is primarily suited for stochastic processes. As discussed also in the original paper (Pompe and Runge, 2011), deterministic processes generate the source entropy that MIT is based on only in the chaotic case due to quantization effects.

References

TM Cover and JA Thomas. *Elements of information theory*. John Wiley & Sons, Hoboken, 2006.

Michael Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 1:233, February 2012. ISSN 0178-8051.

S. Frenzel and B. Pompe. Partial Mutual Information for Coupling Analysis of Multivariate Time Series. *Phys. Rev. Lett.*, 99(20):204101, 2007. ISSN 0031-9007.

Judea Pearl. *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge, 2000. ISBN 0521773628 9780521773621.

Bernd Pompe and Jakob Runge. Momentary information transfer as a coupling measure of time series. *Phys. Rev. E*, 83(5):1–12, May 2011. ISSN 1539-3755.

Jakob Runge, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. Quantifying Causal Coupling Strength: A Lag-specific Measure For Multivariate Time Series Related To Transfer Entropy. *Phys. Rev. E*, 86(6):1–15, October 2012a.

Jakob Runge, Jobst Heitzig, Vladimir Petoukhov, and Jürgen Kurths. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy. *Physical Review Letters*, 108(25):1–4, June 2012b. ISSN 0031-9007.

P. Spirtes, C.N. Glymour, and R. Scheines. *Causation, prediction, and search*, volume 81. The MIT Press, Boston, 2000.

Michael Wibral, Nicolae Pampu, Viola Priesemann, Felix Siebenhühner, Hannes Seiwert, Michael Lindner, Joseph T. Lizier, and Raul Vicente. Measuring information-transfer delays. *PloS one*, 8(2), 2013.